Dissonance and Consistency according to Shackle and Shafer

Isaac Levi

Columbia University

R.A.Fisher introduced the fiducial argument as a means for obtaining something from nothing. He thought that on some occasions it was legitimate to obtain a posterior probability distribution over a range of simple statistical hypotheses without commitment to a prior distribution [4].

H.Jeffreys thought he could tame Fisher by casting his argument in a Bayesian mold through a derivation of the fiducial posterior from a suitably constructed ignorance prior via Bayes' theorem and conditionalization on the data of experimentation. According to Jeffreys, Fisher was using something to obtain something after all ([7],pp.381-383).

D.V.Lindley furthered the process of taming by specifying allegedly necessary and sufficient conditions for the consistency of fiducial reasoning in contexts of one parameter estimation with Bayesian requirements [14]. I. Hacking exploited Lindley's result to supply his own ingenious method for taming Fisher([5],ch. 9).

T.Seidenfeld has recently demonstrated that Lindley's conditions are insufficient for consistency and that reconstructions like Jeffreys' or Hacking's lead to contradiction ([15], pp. 721-727).

Fisher did not want to be tamed and Seidenfeld's result shows that it is not easy to tame him. The brute fact is that he sought to derive a numerically precise posterior without benefit of a numerically precise prior and perforce contradicted Bayesian precepts.

What does it mean to begin with no prior probability

PSA 1978, Volume 2, pp. 466-477

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192483 Published online by Cambridge University Press

judgement? It could mean that one is committed to making no probability judgement at all. Alternatively it could mean that one is committed to making no numerically precise judgement. I favor the latter construal. To make no prior probability judgement is to refuse to rule out any prior distribution (consonant with the calculus of probabilities and other appropriate constraints) and, hence, to be in a state of indeterminate probability judgement represented by the set of all such distributions ([11] and [13]).

When that happens (but not only when that happens) upper and lower probabilities for each simple statistical hypothesis are set at 0 and 1 respectively.

Consider Hacking's simple example of a coin known to have a .4 or a .6 chance of landing heads on a toss where it is known that the coin will be tossed and land heads or tails on the toss. Under suitable circumstances, we may require any numerically precise probability representation to meet the following requirement: $Q(e_H;h_{.6}) = .6$ and $Q(e_H;h_{.4}) = .4$. But aside from the constraint that every precise representation obey the calculus of probability, no further requirement is imposed. In a state of maximally indeterminate probability judgement, probability judgement is represented by the convex hull of the two distributions given below.

	$h_{.4}\&e_H$	$h_{.4}\&e_T$	$h_{.6}\&e_H$	$h_{.6}\&e_T$
Q_1	0.0	0.0	0.6	0.4
Q_2	0.4	0.6	0.0	0.0

Upper and lower probabilities may be read off of this table. They are [0,1] for $h_{.4}$ and $h_{.6}$ and [.4,.6] for e_H and e_T.

Let f_W assert that a winner results on the toss--i.e., that $h_{.4}\&e_T \vee h_{.6}\&e_H$. The chance of obtaining a winner on a toss is .6 so that the credal probability that f_W should be .6 no matter what Q-function in the set is used; and this is precisely what the table reveals. Similarly, the Q-value assigned f_L asserting that the toss yields a loser must be .4.

This simple example illustrates the sense in which there is no prior for the alternatives $h_{.6}$ and $h_{.4}$. No distribution over these two hypotheses assigning some value x in the range from 0 to 1 and 1-x to the two hypotheses respectively has been ruled out.

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192483 Published online by Cambridge University Press

The view that there is no creatio ex nihilo asserts that if the agent observes the result of the toss and adds e_H to his evidence, his new state of probability judgement is obtainable by computing $Q(h_{.6};e_H)$ for every Q-function not ruled out and introducing the distribution $Q'(h_{.6}) = Q(h_{.6};e_H)$ into the new state of probability judgement.

It follows from this that the posterior state of probability judgement for the rival statistical hypotheses is identical with what it was prior to finding out the results of experimentation. That is to say, from nothing one gets nothing.

Of course, this result was based on conditionalization and Bayes' theorem.

Fisher was never bashful in proclaiming the limited usefulness of Bayes' theorem in inverse inference. And he quite clearly thought one could obtain something from nothing. He must have been prepared to abandon conditionalization. No wonder he is so difficult to tame!

Among students of indeterminate probability, the first who, to my knowledge, formally tinkered with conditionalization is H.E. Kyburg (see Levi [12] for discussion and references). Kyburg's approach seems substantially in the spirit of the enigmatic Fisher. He formulated the principle of direct inference in a manner which permitted the derivation of the fiducial argument from that principle--just as Fisher claimed it should be.

According to Kyburg, if agent X knows that the chance of an R on a trial of kind S is r and knows that a trial of kind S&T has been conducted, then, provided that X does not know whether the chance of an R on a trial of kind S&T is or is not different from the chance of an R on a trial of kind S, the degree of probabilistic belief one should assign the hypothesis that an R occurs ought to be r.

Return to our example. The chance of a winner on a toss is known to be .6. After witnessing the outcome of the toss, the agent knows that the toss is not only a toss of the coin but is a toss yielding heads. The agent does not know whether the chance of a winner on a toss yielding heads is different from the chance of a winner on a toss. So according to Kyburg, one should assign the hypothesis that a winner has occurred the probability value of .6 on the datum e_H. But this is equivalent to assigning $h_{.6}$ a posterior value of .6. For all Q'-functions, $Q'(h_{.6}) = .6$ even though there are many Q-values for $Q(h_{.6};e_H)$ distinct from .6. Conditionalization is violated.

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192483 Published online by Cambridge University Press

A few years after Kyburg, A.P. Dempster ([2] and [3]) suggested an alternative way to deviate from conditionalization. In effect, he proposed taking the set of those distributions for which the condition (in our case e_H) receives maximum probability and applying Bayes' theorem to these functions to obtain a set of posterior distributions.

In our example, this leads to an absurd result. The upper probability for e_H is .6. There is only one distribution in the set according to which it bears that value and the posterior for $h_{.6}$ according to that distribution, Q_2, is equal to 1. Everyone should agree that this entails learning from experience too quickly.

There is a way to escape from this predicament and retain Dempster's rule for conditionalization. Prohibit the set of weighted averages of Q_1 and Q_2 from representing a rational state of probability judgement.

This is exactly what Dempster himself does. Indeed, Dempster imposes constraints which prevent any convex subset of the set of weighted averages of Q_1 and Q_2 from being legitimate unless it is a unit set. More striking yet, if the upper probabilities for the two simple statistical hypotheses $h_{.6}$ and $h_{.4}$ are both 1 (and their lower probabilities 0), then either the lower probabilities for e_H and for e_T are less than .4 or the lower probability for the hypothesis that a winner will occur on a toss will be different from the upper probability which, in turn will be different from .6.

I think this result is disturbing and sufficiently so to discredit Dempster's approach; but I am not concerned to explore Dempster's theory in detail. The crucial point is this: Dempster, like Kyburg and Fisher before him, sought to obtain something from nothing by honorable means. They did so in order to avoid becoming caught in the toils of insufficient reason. We may honor their problems and their efforts to solve them even when we decline to accept their solutions.

Glenn Shafer's book, A Mathematical Theory of Evidence, is, in his own words, a "reinterpretation of Dempster's work" ([17], p.ix). An important contribution of Shafer's effort is the clear account of Dempster's formalism which emerges. Reading Shafer can contribute to understanding Dempster.

But an important caveat must be entered. Reading Shafer is no substitute for reading Dempster if one is seeking to understand Dempster's problem and how he sought to solve it. Dempster wrote a foreward to Shafer's book where he alludes to his interest in fiducial inference. That is the sole reference to fiducial inference in Shafer's book or in his PSA paper [18].

In his paper, Shafer compares Bayesian probability with "Dempsterian" probability. This too is misleading. According to Dempster, a state of probability judgement is characterized by those probability distributions the investigator has not ruled out--where the distributions are familiar classical probability distributions. Dempster assumes that the set of such measures obeys a convexity condition and is the largest enveloped by a given system of specifications of upper and lower probability values. Certain still stronger constraints are imposed.

Under these conditions, one can recover the set of distributions which have not been ruled out by specifying the lower probability function over the algebra in question.

But this does not mean, for Dempster, that the lower probability measure is a measure of the agent's degree of belief. Degrees of belief are indeterminate according to Dempster's scheme. When they are numerically definite, they are probabilistic in the classical sense.

In his book, Shafer explicitly disavows this construal--i.e., Dempster's construal--of Dempster's formalism. Instead, he interprets the lower probability function as a measure of degree of belief in a manner which no longer depends on credal states characterizing a state of suspense between several alternative conflicting numerically precise distributions.

According to probabilists who insist that rational agents adopt credal states or states of probability judgement which rule out all but one distribution, the numerically precise credal state assigns degrees of belief to hypotheses in the algebra. On this view, an agent may have a positive degree of belief in a hypothesis even though he does not believe it. Shafer construes degrees of belief quite explicitly so that X assigns h a positive degree of belief (so that $b(h) > 0$) if and only if X believes that h. In terms of support language, the evidence positively supports h if and only if the evidence supports h. Moreover, an agent may fail to believe h and fail to believe -h. In that case, his degree of belief in both propositions is equal to 0 ([17],p.10). Finally, the degree of disbelief or doubt that h, d(h), is equal to the degree of belief that -h. Upper probability or plausibility is $1 - b(-h) = k(h)$.

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192483 Published online by Cambridge University Press

Thus, although the formalism is similar, Shafer's "Dempsterian" probabilities are really "Shaferian" probabilities.

Even this is misleading; for the first person, to my knowledge, to explore degrees of belief and disbelief in a systematic manner and along the lines Shafer investigates is G.L.S. Shackle [16] in the later 1940's. His measure of potential surprise is a measure of disbelief in terms of which a measure of degrees of belief can be defined as just indicated.

Thus, the real novelty of Shafer's approach is to interpret Dempster's formalism as a formalism regulating Shackle's ideas and to provide for Shackle's theory something Shackle sought but did not find--namely, a method for determining how to modify judgements of surprise due to changes in the data.

Shackle provided a formalism for his theory and it is quite different from Shafer's. But given Shafer's own discussion of Shackle, it seems clear that for Shafer the difference between Shackle and Shafer was not over the interpretation of degrees of belief but concerning how degrees of belief ought to be allocated.

According to Shackle, the set of propositions positively believed should form a consistent and deductively closed set.

Shafer dissents. It is possible for the lower probabilities for h and for -h to be positive. If degrees of belief are positive only when propositions are believed, both h and -h must be believed.

Shafer avoids suggesting that h&-h is to be positively believed. If one thinks that this is enough to avoid recommending inconsistent belief, one will be relieved to note that Shafer avoids inconsistency.

Observe, however, that not even Henry Kyburg who, like Shafer, is perfectly willing to abandon deductive closure on beliefs, permits both h and -h to be believed.

Shafer is perfectly aware of this difference between his theory and Shackle's and is proud of it. He claims his view is more general than Shackle's and takes into account phenomena that Shackle's theory cannot handle.

The phenomena to which Shafer alludes he calls "dissonance". Shafer discusses several sorts of dissonance but the most striking is illustrated by the following passage ([17], pp.84-85):

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192483 Published online by Cambridge University Press

> 4.3 The alibi. A criminal defendant has an alibi. A close friend swears that the defendant was visiting his apartment at the time of the crime. The friend has a good reputation, so the testimony carries some weight in spite of his close friendship with the defendant. Let us suppose that standing alone it would provide a degree of support of 1/10 for the defendant's innocence. But ranged on the other side is a strong body of circumstantial evidence attesting to the defendant's guilt; standing alone it would provide a degree of support of 9/10 for his guilt. What degrees of support does the combined evidence provide for the defendant's guilt and innocence?

Shafer then procedes to analyze the problem according to his principles for "combining" evidence. It is a consequence of his principles very clearly and deliberately built into them that the combined evidence should positively support the hypothesis of innocence to some degree and the hypothesis of guilt to some degree. That is to say, the evidence taken together should warrant a positive degree of belief that the defendant is innocent and another higher positive degree of belief that he is guilty.

According to Shafer, examples like this illustrate conflict or dissonance in the evidence even when the propositions reporting the evidence are perfectly consistent. In commenting on Shackle and L.J. Cohen (whose measure of inductive support he mistakenly takes to be a Shackle measure), Shafer writes as follows ([17], pp.225-226):

> It is easy to share the desire of these scholars to ban the appearance of conflict from our assessment of evidence and our allocation of belief. But in the light of what we have learned, the ambition of doing so must be deemed unrealistic. The occurrence of outright conflict in our evidence should and does discomfit us; it prompts us to reexamine both our evidence and the assumptions that underlie our frame of discernment with a view to removing that dissonance. But this effort does not always bear fruit--at least not quickly. And using all the evidence often means using evidence that is embarrassingly conflicting.

In the coin example, suppose the coin had been tossed twice and had landed heads once and tails once. Shafer acknowledges dissonance in cases like this. Such dissonance should "discomfit" us on his view and lead us to reexamine the data or modify the range of rival hypotheses being entertained. But surely it would be neurotic to be discomfited in the case of the coin example simply because part of the

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192483 Published online by Cambridge University Press

data when taken alone positively supports h and the other part when taken alone positively supports h'.[6] No one, including Shackle, would or should find this[4] "discomfiting". Nor should anyone find it upsetting that part of the evidence points to the defendant's guilt and another part to his innocence.

What Shackle would find upsetting is the recommendation to believe both contraries when the data are pooled or to take both contraries to be supported by the total evidence.

It is one thing to point out that part of the evidence points to the defendant's innocence and another part to his guilt. From this it does not follow that the total evidence partially supports his innocence and partially supports his guilt.

To be sure, that is a consequence of the phenomenon of dissonance--which everyone agrees can and does occur--and Shafer's rule for combining data. To charge those who deny that the total evidence should support both contraries in such cases with a lack of realism is to beg the fundamental point under dispute--to wit, whether the rule for combining data is acceptable.

To my knowledge, Shafer has offered no argument in favor of his rule for combining data except the appeal to dissonance which, as we have just seen, is no argument at all.

To those convinced beforehand that a body of beliefs should be deductively consistent and closed, the phenomenon of dissonance points decisively to the conclusion that Shafer's rule of combination must be untenable. For such thinkers (I include myself in their number), "realism" undermines Shafer's enterprise and does so decisively.

To those who refuse to endorse deductive closure perhaps because of the reasons adduced by Kyburg but who cling to the idea that no one should believe simultaneously both h and -h, Shafer's rule of combination should also seem untenable.

Of course, those who deny the intelligibility or importance of notions of qualitative belief and rest content with Bayesian probabilities or, perhaps, states of credal probability judgement representable by sets of distributions will declare a plague on all other houses including Shafer's --at least as long as Shafer is taken at his word and is understood to be interpreting Dempster's lower probabilities as degrees of belief in the sense in which positive belief is belief construed qualitatively.

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192483 Published online by Cambridge University Press

Thus, Shafer stands opposed to almost everybody. Such daring would be admirable were it accompanied by an explanation of the problems which Shafer thinks can be solved better by adopting his approach rather than one of the alternatives or by an explanation of what he means of degrees of belief or support which shows far more clearly than he has done that his notion is distinct from others and is intended for other applications than those usually envisaged.

What Shafer has done instead has been to pour Shackle's notion of degrees of belief and disbelief into Dempster's formalism so as to require modifying Shackle's formalism and has excused himself by appealing to dissonance which, as I have indicated, presupposes the principles for combining evidence that mandate the questionable deviations from Shackle's view in a question begging manner.

Perhaps, however, we should credit Shafer with addressing a problem rarely confronted--namely how to determine degrees of belief on combined evidence. Even if his own proposal is in trouble, surely his challenge ought to provoke us to do better.

But Shackle also worried about a similar question and he proposed his own answer which he then had to modify and which, even after modification, remained a shambles. So the credit for raising the problem should go to Shackle.

In discussing Shackle's troubles, I suggested a reconstruction of his theory which related his measures of surprise and belief to caution dependent families of rules for acceptance. I claimed that by means of such rules one could give a role to Shackle's measures as indices of weight of evidence or argument in the sense of Keynes. I also stated that the correct way to approach revision of assignments of degrees of belief or support with new data would be through application of caution dependent families of rules relative to the various bodies of data ([8];[9], chapters 8 and 9; and [10]).

Thus, for a full decade before Shafer published his book, we have had an account of degrees of belief and disbelief and, incidentally, of plausibility as well, which is accompanied by an account of how such degrees are to be modified with alterations of the data. This account allows for the phenomenon of dissonance but in a manner which does not entail the "discomfiture" of allowing belief in contrary hypotheses.

From Shafer's point of view as stated in his publications, the sole liability would seem to be that this proposal fails to supply an interpretation of Dempster's formalism.

I think Dempster has raised interesting questions of his own. Whatever difficulties there may be with his approach, the merits of his work can stand scrutiny without the reinterpretation of his formalism Shafer has offered us.

In the 1960's, I pointed out that potential surprise (degrees of disbelief) and degrees of belief do not appraise hypotheses in a manner rival to probabilistic modes but are complementary rendering different services in deliberation and inquiry than either numerically precise Bayesian probabilities or indeterminate probability judgements.

In my opinion, Shackle overstated the conflict between his measures of surprise and probabilistic measures as epistemic measures of uncertainty. In his paper [18], Shafer contrasts Bayesian probabilities with "Dempsterian" or, more accurately, "Shaferian" probabilities. My contention is that there is no rivalry--or, at any rate, there should not be.

Long ago Keynes distinguished between probability as balance of argument and something else he called "weight" of argument. It is clear from his discussion that he thought both epistemic modes of assessment had their proper roles to play in inquiry and deliberation. I wish to echo his sentiments.

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192483 Published online by Cambridge University Press

References

[1] Carter, C.F. and Ford, J.L. (eds.). Uncertainty and Expectation in Economics. Oxford: Blackwell, 1972.

[2] Dempster, A.P. "Upper and Lower Probabilities Induced by a Multivalued Mapping." Annals of Mathematical Statistics 38 (1967): 325-339.

[3] -------------. "A Generalization of Bayesian Inference." Journal of the Royal Statistical Society, Series B 30(1969): 205-247.

[4] Fisher, R.A. "Inverse Probability." Proceedings of the Cambridge Philosophical Society 26(1930): 528-535.

[5] Hacking, I. Logic of Statistical Inference. Cambridge: Cambridge University Press, 1965.

[6] Hooker, C.A., Leach, J.J., McLennen, E.F. (eds.). Foundations and Applications of Decision Theory, Volume 1. (University of Western Ontario Series in The Philosophy of Science, Volume 13). Dordrecht: Reidel, 1978.

[7] Jeffreys, H. Theory of Probability. 3rd ed. Oxford: Clarendon Press, 1961.

[8] Levi, I. "On Potential Surprise." Ratio 7(1966): 107-129.

[9] -------. Gambling with Truth. New York: Knopf, 1967. Reprinted by MIT Press, 1973.

[10] -------. "Potential Surprise in the Context of Inquiry." In [1]. Pages 213-236.

[11] -------. "On Indeterminate Probability." Journal of Philosophy 71(1974): 391-418. Reprinted in [6] with addenda. Pages 233-261.

[12] -------. "Direct Inference." Journal of Philosophy 74(1977): 5-29.

[13] -------. "Irrelevance." In [6]. Pages 263-273.

[14] Lindley, D.V. "Fiducial Distributions and Bayes' Theorem." Journal of the Royal Statistical Society Series B 20(1958): 102-107.

[15] Seidenfeld, T. "Direct Inference and Inverse Inference." Journal of Philosophy 75(1978): 709-730.

[16] Shackle, G.L.S. "Expectations in Economics." Cambridge: Cambridge University Press, 1949.

[17] Shafer, G. A Mathematical Theory of Evidence. Princeton, N.J.: Princeton University Press, 1975.

[18] ----------. "Two Theories of Probability." In PSA 1978, Volume 2. Edited by P.D. Asquith and I. Hacking. East Lansing, Michigan: Philosophy of Science Association, 1981. Pages 441-465.

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192483 Published online by Cambridge University Press